Development of Pedotransfer Functions and Related Computer Programs

M.G. Schaap, F.J. Leij and M.Th. van Genuchten
U.S. Salinity Laboratory, Riverside, CA

Concern about the quality of soil and water resources has motivated the development of increasingly sophisticated models describing water, heat and solute movement in unsaturated soils. Models for variably-saturated flow and heat transport are also needed to simulate the atmospheric boundary layer for predictions of regional or global climate change, and to optimally interpret or improve the utility of remotely sensed data. Unsaturated soil hydraulic models are generally based on numerical solutions of the Richards equation. Such solutions rely on expressions for the soil water retention curve, $\theta(h)$, which relates the volumetric soil water content with the soil water pressure head, h, and the unsaturated hydraulic conductivity function, $K(h)$. Popular expressions for $\theta(h)$ include the Brooks-Corey (1963) and van Genuchten (1980) equations. Functions by Gardner (1958) and van Genuchten (1980) are frequently used for $K(h)$. Direct measurements of θ(h) and especially $K(h)$ are difficult and expensive. Lack of knowledge of these properties is hence often a weak link in modeling efforts. As an alternative, pedotransfer functions (PTFs) can be used to predict hydraulic properties from texture, bulk density or other soil taxonomic data. Many PTFs have been published; studies concerning their applicability have been carried out by Tietje and Tapkenhinrichs (1993) and Schaap et al. (1998), among others.

When using PTFs three practical questions arise. One question deals with the type of surrogate data that are best used to estimate soil hydraulic properties. A related issue is to extract the most information from data that are already available. Second, because many PTFs exist, which one provides the most accurate predictions? And third, since the PTF predictions are *estimates*, what is the associated uncertainty?

This project addresses these questions by developing hierarchical PTFs that are able to use different levels of input data. The approach permits predictions even if only very limited information is available (e.g., only the soil textural class), but also provides possibilities to improve the predictions when more information becomes available later (e.g., textural distribution, bulk density). We used neural networks to obtain the most accurate PTFs. Studies by Pachepsky et al. (1996), Schaap and Bouten (1996), and Schaap et al. (1998) have shown that neural network results are often better than more traditional regression methods. By combining neural networks with the bootstrap method (Efron and Tibshirani, 1993) one can calculate a probabilistic distribution of the hydraulic properties, and thus provide confidence intervals.

In the following we provide a general overview of a hierarchical prediction of the saturated hydraulic conductivity (K_s) and the hydraulic parameters θ_r, θ_s, α and n in van Genuchten's (1980) water retention curve:

$$S_e = \left[1 + (\alpha h)^n\right]^{1/n-1} \qquad (1)$$

where

$$S_e = (\theta - \theta_r)/(\theta_s - \theta_r) \quad (2)$$

A similar hierarchical approach, including estimates of confidence intervals, is currently being developed for unsaturated hydraulic conductivity.

We also provide here some interesting results for the statistical pore-size model of van Genuchten (1980) and Mualem (1976) to predict the conductivity according to

$$K(S_e) = K_s S_e^L [1 - \{1 - S_e^{n/(n-1)}\}^{1-1/n}]^2 \quad (3)$$

where S_e is computed with Eq. 1. We will concentrate on the estimation of the exponent L.

Hierarchical Approach and Confidence Intervals for $\theta(h)$ *and* K_s

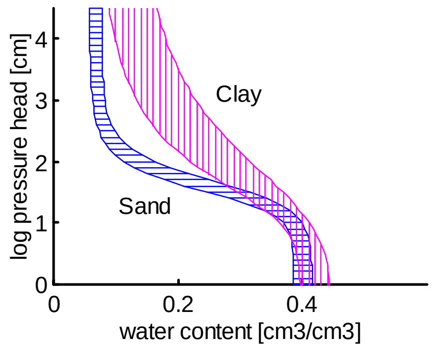
We considered five levels of input data (cf. Table 1):

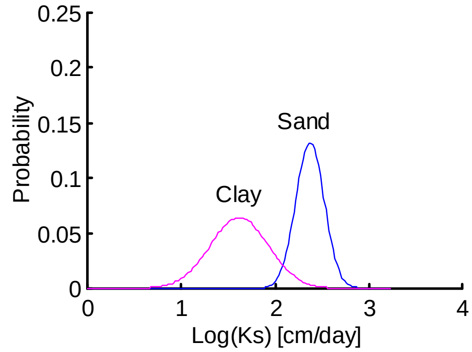
1) Soil Textural class (TXT),
2) Percentages of sand, silt and clay (SSC),
3) As in 2) but with addition of bulk density (SSCBD),
4) As in 3) but with addition of a water retention point at 33 kPa (SSCBD+θ_{33}), and
5) As in 4) but with addition of a water retention point at 10 kPa (SSCBD+$\theta_{10}\theta_{33}$).

Although the use of water retention points as input in level 4 and 5 may seem unusual for predicting water retention parameters, previous studies have shown that one or two points can improve the predictability of water retention curves significantly (e.g. Schaap et al., 1998). Furthermore, databases like the large NRCS database often include some information about water retention.

Table 1 shows the accuracy of PTFs as root mean square residuals (RMSR) for both water retention and K_s.. The RMSR values for retention and K_s decreased when more input variables are used, indicating an improvement in the prediction. The smallest RMSR values were obtained by adding bulk density and one retention point (the SSCBD and SSCBD+θ_{33} models). Using two retention points is not as effective. Using texture percentages instead of textural classes also did not substantially improve the prediction of water retention and K_s.

Comparison of the five hierarchical models with published pedotransfer functions (also given in Table 1) shows that the neural network models provide somewhat better predictions than the previously published explicit water retention PTFs. The neural network models for K_s are clearly much better than the published PTFs. The best neural network models were able to predict K_s within half an order of magnitude.





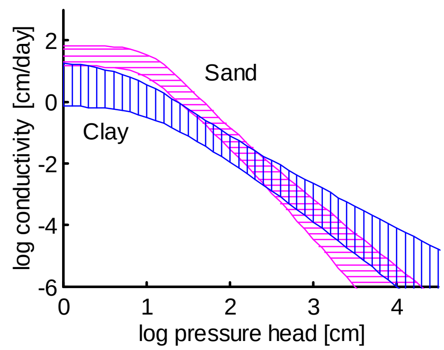


Figure 1. *(Top Left) Ten and 90% uncertainty intervals for water retention as predicted by the SSCBDθ_{33} model for the water retention curve of a loamy sand*

Figure 2. *(Top Right) Uncertainty for K_s as predicted by the SSCBDθ_{33} model for a loamy sand and a clay.*

Figure 3. *(Bottom) Ten and 90% uncertainty intervals for the unsaturated hydraulic conductivity (Gardner's function) as predicted by the SSCBDθ_{33} model for a loamy sand.*

Figures 1 through 3 show for both a loamy sand and a clay soil the uncertainty in, respectively, the retention function (Eq. 1), the saturated hydraulic conductivity, K_s, and Gardner's $K(h)$ equation (1958). The graphs were generated using a combined bootstrap-neural network approach and level 4 as input (SSCBDθ_{33}). For water retention and the Gardner function, 10 and 90% percentiles of the variability among the submodels are shown, while for K_s we plotted the entire probability distribution. Directly apparent from the figures is the larger uncertainty for the clay as compared to the loamy sand. This is caused by the smaller number of fine-textured soils in the data set relative to coarse-textured soils. In general, the uncertainty estimates increased for samples that were less common in the calibration data sets.

__Table 1.__ Results of neural networks and previously published PTFs om terms of root mean square residuals (Sa: sand, Si: silt, Cl: Clay, OC: organic carbon, NA: not available).

Model	Input	-------------RMSR------------	
		Retention $cm^3\ cm^{-3}$	K_s $\log(\mathrm{cm\ day}^{-1})$
Neural network PTF			
TXT	Textural class	0.107	0.627
SSC	Sa, Si, Cl%	0.104	0.602
SSCBD	Same + bulk density	0.087	0.533
SSCBD+θ_{33}	SSCBD + 33 kPa retention	0.060	0.451
SSCBD+$\theta_{10}\theta_{33}$	Same + 10 kPa retention	0.058	0.448
Published PTF			
Rawls et al. (1985)	Sa, Cl, porosity	0.101	NA
Cosby et al. (1984)	Sa or Cl	0.111	0.746
Vereecken et al. (1989, 1990)	Sa, Cl, BD, OC	0.098	0.934
Brakensiek et al. (1984)	Sa, Cl, porosity	NA	0.791
Saxton et al. (1986)	Sa, Cl, porosity	NA	0.761
Ahuja et al. (1989)	Porosity, 33 kPa retention	NA	0.822

Unsaturated Hydraulic Conductivity

A major objective of our analysis was to accurately estimate the exponent L in Eq. (3). Mualem (1976) initially found an average value of 0.5 for a dataset containing some 50 disturbed and undisturbed samples. Later studies found that the value of L can greatly vary, possibly as a function of soil type. Using 250 samples of the UNSODA international soil-hydraulic database (Leij et al., 1997) we found that a value of 0.5 leads to higher fitting errors than an near-optimal value of -1.0 (Figure 4). We also found that L was related to the parameter n in Eq. (1) (see Figure 5). Root mean square residuals of the predicted unsaturated hydraulic conductivity curves were 1.4 and 1.3 log(cm/day) when L was set to 0.5 and -1.0 respectively. When L was considered to be a variable, the RMSR decreased to 1.1 log(cm/day). Similar results were obtained for the empirical $K(h)$ function by Gardner (1958).

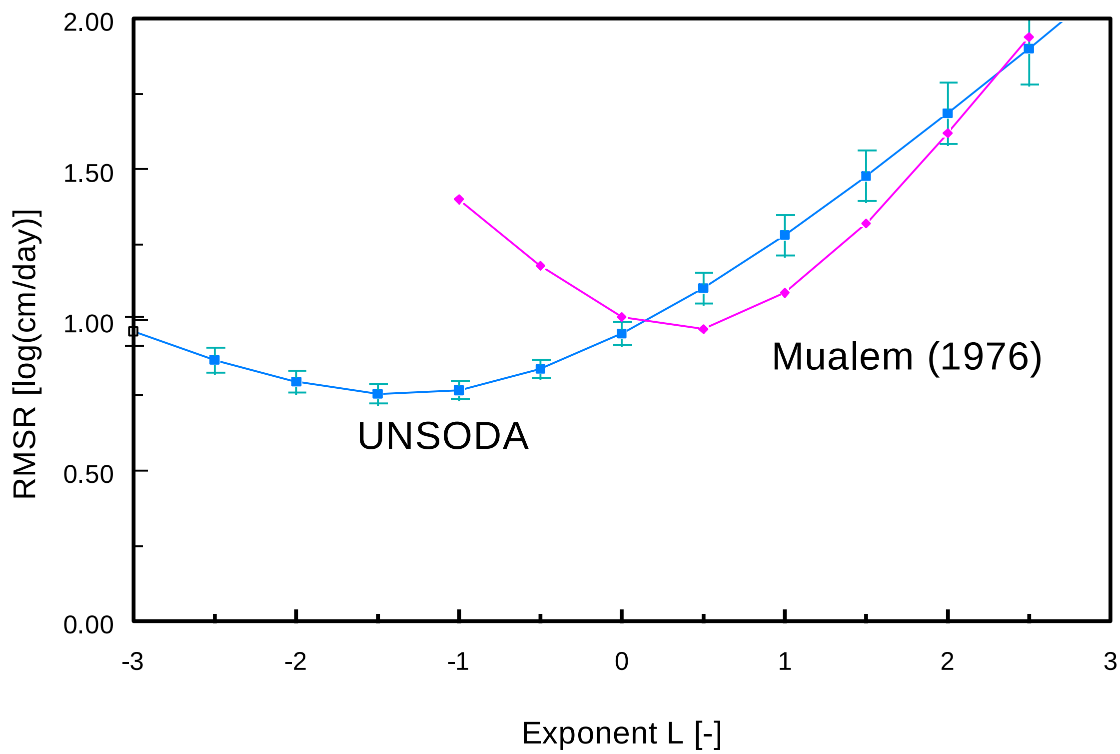


Figure 4. *Optimal values for the exponent L (Eq. 3) according to our studies with the UNSODA data base and as found by Mualem (1976).*

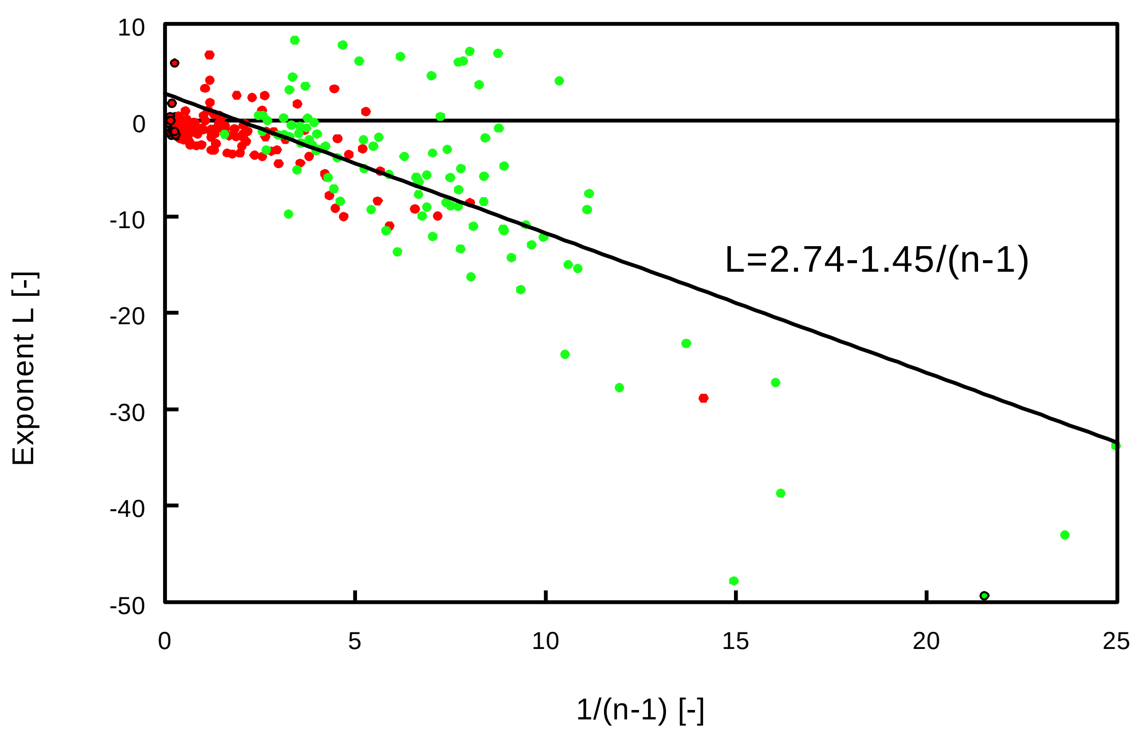


Figure 5. *Dependence of the exponent L (Eq. 3) on parameter n in Equation 1.*

Conclusion

We created an hierarchical system of neural network models that predicts soil hydraulic properties from different levels of input data. The bootstrap-neural network approach provided relatively accurate predictions as well as uncertainty estimates. These results are needed to generate uncertainty estimates of water and solute transport processes using, for example, Monte Carlo simulations. Uncertainty estimates also give guidance as to where one could further improve models by collecting more samples for input combinations which have relatively large uncertainties. We believe that our hierarchical approach is not only attractive because of improved accuracy, but also because neural network models can be chosen to match a particular set of available data. Because the hierarchical neural network models were calibrated on the same data set, predictions of the different models will be consistent. Work is currently underway to implement the neural network models in a user-friendly software package.

Publications resulting from this project

Schaap, M.G., F.J. Leij and M. Th. van Genuchten. 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. Accepted by Soil Sci. Am. J.

Schaap, M.G. and F.J. Leij. 1998. Using Neural Networks to predict soil water retention and soil hydraulic conductivity. Accepted by Soil Technology.

Schaap, M.G. and F.J. Leij. 1998. Calibration, validation and cross-validation of neural network pedotransfer functions. Submitted to Soil Science.

Schaap, M.G., F.J. Leij and M. Th. van Genuchten. 1998. A bootstrap-neural network approach to predict soil hydraulic parameters. To appear in the proceedings of the workshop: Characterization and measurement of the hydraulic properties of unsaturated porous media, Riverside, CA, Oct 22-24, 1997.

Schaap, M.G. and F.J. Leij. 1998. Comparison of methods to predict unsaturated hydraulic conductivity. To be submitted to Water Resour. Res.

Other references

Brooks, R.H., and A.T. Corey. 1964. Hydraulic properties of porous media. Hydrol. paper 3, Colorado State Univ., Fort Collins, CO.

Efron, B., and R.J. Tibshirani. 1993. An Introduction to the bootstrap. Monographs on statistics and applied probability. Chapman and Hall, New York, NY.

Gardner, W.R. 1958. Some steady-state solutions of the unsaturated moisture flow equations with application to evaporation from a water table. Soil Sci. 85: 228-232.

Leij, F.J., W.J. Alves, M.Th. van Genuchten, and J.R. Williams. 1996. The unsoda unsaturated soil hydraulic database, version 1.0. EPA report EPA/600/R-96/095, EPA National Risk Management Laboratory, G-72, Cincinnati, OH, USA. (available at: http://www.epa.gov/ada/models.html)

Mualem, Y. 1976. A new model predicting the hydraulic conductivity of unsaturated porous media. Water Resour. Res. 12: 513-522.

Pachepsky, Ya. A., D. Timlin, and G. Varallyay. 1996. Artificial neural networks to estimate soil water retention from easily measurable data. Soil Sci. Soc. Am. J. 60: 727-733.

Schaap, M.G., and W. Bouten 1996. Modelling water retention curves of sandy soils using neural networks. Water Resour. Res. 32: 3033-3040.

Tietje, O., and M. Tapkenhinrichs. 1993. Evaluation of pedotransfer functions. Soil Sci. Soc. Am. J. 57:1088-1095.

van Genuchten, M.Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, Soil Sci. Am. J. 44: 892-898.